

### REMARKS

1. Applicant has appended hereto proposed new formal drawings.
2. Applicant has appended hereto a new declaration.
3. Applicant has amended claim 1 to include the restrictions of claim 2, and canceled claim 2. The Examiner has indicated that claims 2-11 contain allowable material

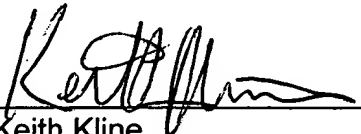
In light of the above amendments and remarks, Applicant now asserts that all of the grounds for rejection have been traversed or overcome by amendment, and that all of the present claims are in condition for immediate allowance. Applicant therefore requests reconsideration of the objections and rejections, and solicits allowance of the present claims at an early date.

Thank you for your consideration.

Respectfully submitted,

Date: June 17, 2004

Registration No: 32,737  
Telephone: (408) 778-3440

  
\_\_\_\_\_  
Keith Kline

PRO-TECHTOR INTERNATIONAL  
20775 Norada Court  
Saratoga, CA 95070-3018

S/981010

uv - replace "address bits" w/ "address"  
4 11 statistical w/ "stat. cal"

## STRUCTURE AND METHOD OF REPAIRING SDRAM BY GENERATING SLICING TABLE OF FAULT DISTRIBUTION

### Field of Invention

This invention is generally related to the structure and method for mending or repairing SDRAM. More specifically, it relates to the generation of Slicing Table of Fault Distribution, and encoding the repair bit according the Slicing Table of Fault Distribution. By pointing to the remapping bit, the access action occurs at the remapping bit so as to maintain the quality of the SDRAM product.

### Background of Invention

Any for profit SDRAM producer would know that quality of the SDRAM product is a key factor in determining the profit. Therefore, a majority of the IC producers have invested money in reducing the number of defective product, including incentives for workers, technicians and engineers to find ways to enhance the quality of the IC product. After analyzing the production volume of the IC producers for the past 25 years, a 1991 publication in the form of a VLSI dissertation stated that an IC producer generally invests at least 600 to 700 million dollars to increase the qualification rate to at least 85 percent. Moreover according to a 1991 publication by the Semiconductor Industry Union entitled the "International Technology Roadmap for Semiconductors in Defect Reduction", it is stated that the qualification rate is limited to between 85 to 95 percent for the IC producer regardless of the debasing deficiency generated by the product-processing or the equipments, as shown in Fig. 1.

To reduce the waste cost of the 5 to 15 percent defective product, SDRAM producers generally undertake a repair program to mend or repair the 5 to 15 percent defective product, so as to raise the qualification rate. A few of the common repair methods adopted by the SDRAM producer are discussed hereinbelow:

#### (1) Logical Method

As shown in Fig. 2, the various features of this method include the testing step and patch step. The testing-step is to test whether the memory 100 has defective cells 110, and if it has, record the address bits of the defective cells 110. The patch step is to use comparator 120 to compare the recorded defective address bits 130 in the testing-step with the address bits of access data transferred by the CPU, and if the address bits of access data is in the record, it will re-map the address bits to the patch-memory 140 (usually SRAM). By using one of the cells in patch-memory to replace the defective cell in the testing-memory, it would cure the defective product. However, the shortcomings of this repair method are:

1. The address bits signal, which is transferred by the CPU, is not only presented to the testing-memory 110, but also to the patch-memory 140. So this address bits signal must have enough fan-out power or it can be mistakenly distinguished.

2. Since the repair process is slow the comparator 120 must compare the address bits one by one, and it can take a long time to patch the

memory page.

3. There must be a colossal scale and high ability comparator 120 ~~to be presented~~, which is very expensive.

## (2) Spare Fault-tolerant Method

As shown in Fig.3, the main feature is the repair process of the wafer test. Memory 200 will make spare array elements 220 while making the wafer. During the test, CPU sends the row address ~~bits~~ 230 to the cell of memory 200 first, then send the column address ~~bits~~ to test every memory cell after a proper TRCD. Once finding the defective cells 210, it replaces the defective cells 210 with spare array elements 220 in a radiate mode to make the defective product effective. The shortcomings of this repair method are:

1. The repair only occurs during the wafer test. If the wafer is cut or sealed, the method would fail.

2. It takes 1.5 seconds to repair every die, which is too long ~~of a repair time~~.

3. It increases the cost of repair in the radiate mode.

The above-described conventional repair methods, especially the latter one, greatly increase the investment cost for the SDRAM producer.

## Summary of Invention

The invention is directed to a structure and method for repairing a memory defect in SDRAM to reduce cost and increase profit for the SDRAM producer.

In order to achieve the above object, the invention generates a Slicing Table of Fault Distribution during booting or memory-testing. The particular steps are as follows: dividing the SDRAM into memory pages of a proper size; using 4k as the criteria according to the statistic data of the memory fault distribution, and a range of  $4k \cdot 2^n$  as the partition unit; for the possible defects in every 8k, 16k, 32k, and 128k memory, posting the number of the tolerate-mistake memory pages, arranging and preserving the number of the memory and the quantity of the comparator correspond to the address ~~bits~~ of the incorrect memory pages.

The present invention also uses the Slicing Table of Fault Distribution to repair SDRAM. The particular steps are as follows:

- a. booting the system;
- b. executing memory-test to find the address and distribution of the fault in the memory chip;
- c. if the fault distribution is in the planned range in advance, establishing a slicing table of fault distribution ;
- d. if the fault distribution is concentrated in a certain range, establishing a slicing table of fault distribution and posting to the address limiter;
- e. inputting the address ~~bits~~ into the address limiter;
- f. using the comparator arrays to check whether the input address ~~bits~~ is in  $a$  different address ~~bits~~ ranges according to the slicing table of fault distribution;
- g. if there are no faults, mapping the input address ~~bits~~ to the SDRAM and reading data from that address ~~bits~~;
- h. if there are some faults, generating a fault enable signal so that the system

will not read any data at the original input address ~~bits~~ and generate a signal of fault transferred to the encoder;

- i. using the encoder to set up a repair address ~~bits~~ for the fault;
- j. using the repair memory to point to a new remapping address ~~bits~~ according to the repair address ~~bits~~;
- k. replacing the original input address ~~bits~~ with the remapping address ~~bits~~ so that data will be read out at the remapping address ~~bits~~ in the SDRAM.

Additionally, a structure is presented according to the above-described methods, and includes at least the followings: a plurality of Address Limiters, a plurality of correspond memories, a plurality of comparator arrays, an encoder, a repair memory and a multiplexer.

### **Brief Description of Drawings**

Fig.1 illustrates the qualification rate statistics of the memory of the IC manufacture factories recently;

Fig. 2 illustrates the logical repair method for the memory;

Fig. 3 illustrates the spare fault-tolerant method for the memory;

Fig. 4 illustrates the normal fault distribution of the memory according to the statistics data;

Fig. 5 illustrates the Slicing Table of Fault Distribution in the embodiment;

Fig. 6 illustrates the outline map of the structure in the embodiment;

Fig. 7 illustrates that the graphical chart of fault distribution in the figure 4 drifts from central to left; and

Fig. 8 is a flow chart of this invention.

### **Detailed Description of Invention**

A preferred embodiment accompanied with the drawings is explained herein below in detail.

From the statistics<sup>al</sup> data as shown in Fig.4, the quantity of fault at, for instance, 8k, 16k, 32k, or 128k memory is used to establish the Slice Table of Fault Distribution. By ~~not every value or under value~~ <sup>are working defect memory</sup> the number of the defective memory pages, all the resources ~~is~~ <sup>are</sup> efficiently used to ~~a maximum degree~~. In this invention, the empirical value is preserved in the memory chip, so that when the system is booted and the SDRAM pages are checked, the Slicing Table of Fault Distribution is established.

If there is a 2G SDRAM, then every memory page has 4k bytes for a total of 521k memory pages. Since the statistical data of all defective memory pages in the 2G SDRAM and the number of the defective memory pages according to every 8k, 16k, 32k, 128k under ~~the~~ normal conditions are known, the Slicing Table of Fault Distribution would be established according to these data, and ~~filled~~ <sup>set up</sup> the associate memory and correspond comparator, and preserved ~~the~~ <sup>by</sup> the remapping address ~~bits~~ <sup>bits</sup> of the defective memory pages.

Although executing more comparators at the same time would accelerate the process of locating and repairing defective pages, it will also cost more in the surplus relevant memory (i.e., SDRAM) and comparator. Therefore, using more than 16

relevant memories and comparators <sup>has</sup> have not been considered due to cost consideration.

Fig.5 illustrates an example of contents of the Slicing Table of Fault Distribution. In particular, for the statistical data of memory fault distribution, 4k is used as the criteria, and every  $4k \cdot 2^n$  range is used as the partition unit ( $n$  is the integer more than or equals 0) to divide the memory pages. In this example, every 8k, 16k, 32k, and 128k are used as the partition unit, and different ranges can be utilized based on different conditions. In all these ranges, the number of the tolerant fault memory pages, and the number of the associate memories for preserving the address ~~bits~~ of the fault memory pages are set up as follows:

1. Two fault memory pages (memory cells) are permitted in every 8k memory pages, and two associate memories are arranged to preserve the address ~~bits~~ of fault memory pages found in the range of 8k memory pages, and two comparators to compare the address ~~bits~~ stored in the associate memories with ~~the~~ input address offset ~~bits~~.

2. Two fault memory pages (memory cells) are permitted in every 16k memory pages, and two associate memories are arranged to preserve the address ~~bits~~ of fault memory pages found in the range of 16k memory pages, two comparators to compare the address ~~bits~~ stored in the associate memories with ~~the~~ input address offset ~~bits~~.

3. Two fault memory pages (memory cells) are permitted in every 32k memory pages, and two associate memories are arranged to preserve the address ~~bits~~ of fault memory pages found in the range of 32k memory pages, two comparators to compare the address ~~bits~~ stored in the associate memories with input address offset ~~bits~~.

4. Four fault memory pages (memory cells) are permitted in every 128k memory pages, and four associate memories are arranged to preserve the address ~~bits~~ of fault memory pages found in the range of 128k memory pages, four comparators to compare the address ~~bits~~ stored in the associate memories with input address offset ~~bits~~.

Since every memory page is 4k, there are a total of 512k memory pages for a 2G SDRAM.

If the partition unit is a 8k memory page (i.e., two 4k memory page), it will be divided into sixty-four 8k memory pages. At the same time, 2 fault memory pages are tolerant in every 8k memory pages, so that overall 128 fault memory pages are tolerant in this partition range.

If the partition unit is a 16k memory page (i.e., four 4k memory page), it will be divided into thirty two 16k memory pages. At the same time, 2 fault memory pages are tolerant in every 16k memory pages, so that overall 64 fault memory pages are tolerant in this partition range.

If the partition unit is a 32k memory page (i.e., eight 4k memory page), it will be divided into sixteen 32k memory pages. At the same time, 2 fault memory pages are tolerant in every 32k memory pages, so that overall 32 fault memory pages are tolerant in this partition range.

If the partition unit is a 128k memory page (i.e., thirty-two 4k memory page), it will be divided into four 128k memory pages. At the same time 4 fault memory pages

are tolerant in every 128k memory pages, so that overall 16 fault memory pages are tolerant in this partition range.

Based on the above-description, 240 fault memory pages ( $128+64+32+16=240$ ) are permitted in a 2G SDRAM.

In the following, an example is used to explain the actual work (from the input address ~~bits~~ to the output address ~~bits~~) in the memory chip. As shown in Fig. 6, the structure of the memory chip includes: address limiters 310, associate memories 320, comparator arrays 330, encoder 340, repair memory 350 and multiplexer 360.

Referring to the address limiter 310, to the Slicing Table of Fault Distribution, when the number of faults exceeds the limit of one specific partition, the address ~~bits~~ of the block is restricted, and this fault is cured by using the limit of fault of other partition with a lower fault rate. Fig.4 illustrates that the normal fault distribution area of the memory is based on the statistics<sup>31</sup> data. However when the SDRAM has a higher serious fault rate at the lower address ~~bits~~ range, the graphical chart of fault distribution drifts leftward from a central position as shown in Fig.7. The number of the actual fault pages (cells) at the lower address ~~bits~~ range will exceed the number of the tolerant fault pages in the partition area.

In the example as shown in Fig.7, there are about less than ten fault pages in the range of 32k memory pages. But according to the criteria of the tolerant fault pages in a partition, i.e., 2 fault pages in 8k, 4 fault pages in 16k, 6 fault pages in 32k and maximum 10 fault pages in 128k, one must use other 128k partition to overlay these surplus fault pages. Although the distribution of the tolerant faults can be arranged, the total of the tolerant fault is not to be changed. In the other word, if a partition has more fault than the limit so as to not change the total tolerant fault, then the fault limit of other partition has to be sacrificed to make up for this special partition, and this method is called address limit of the memory. When the computer system opens up, the system finds the memory address partition having the maximum fault after testing the memory, and the fault limit of ~~other several~~ partitions is sacrificed to make up this special partition so that the fault overlays every flexible partition.

Referring to the associate memory 320, it is used to preserve the address ~~bits~~ of the fault memory pages. The quantity is associated with the tolerant fault for the partition. As shown in Fig.5, in a 2G bit ( $\sim 2^{31}$  bit) SDRAM (address ~~bits~~: A30:A0), every memory page is 4k ( $\sim 2^{12}$  bit) (address ~~bits~~: A11:A0), so that there are 512k ( $\sim 2^{19}$  bit) memory pages totally (address ~~bits~~: A30:A12). According to the example Slicing Table of Fault Distribution as shown in Fig.5, two fault memory pages are tolerant in every 8k memory pages and two associate memories are arranged to preserve the address ~~bits~~ of fault memory pages. As the two and four associate memories are increased in surplus to preserve the address ~~bits~~ of fault memory pages in every 16k, 32k and 128k memory pages. A bit is appended to validate whether there is any fault (0 is for no fault, 1 is for having fault), so that the size of every associate memory is  $(2G/32M) \times (\text{the address offset} + 1)$  bit for the range of 8k ( $\sim 2^{25}$  bit) memory. This address offset is A30: A25.

Comparator array 330 compares the input address ~~bits~~ 380, the address ~~bits~~ selected by the address limiters 310, and the output address bits of the associate

memory 320, then judges whether they are of different ranges of the memory address ~~bits~~ and determines whether there is any fault. When the input address ~~bits~~ 380 wanted by the SDRAM is transferred into the memory chip, the address ~~bits~~ will be selected by the memory address limits 310. In order to be selected, the input address ~~bits~~ 380 and the output address ~~bits~~ of the associate memory 320 are checked by the associate memory 330 at the same time so as to determine whether they are of different memory address ~~bits~~ ranges. Delay occurs by association, but the memory works at the same association, so that there is no efficacy loss. If no defective page is found, data are to be read in the normal state at the address ~~bits~~; if there are defective pages, a signal of fault is transferred to the encoder 340.

Referring to the encoder 340, after receiving the signal of fault, the encoder 340 will generate remapping address ~~bits~~ correspond to the repair memory 350. The repair memory 350 includes the information of the remapping address ~~bits~~. The repair address ~~bits~~ is generated by a special encoding method, and includes two type of data, one of which is the tolerant faults in the memory page range, and the other one is the partition of the input address ~~bits~~.

Referring to the repair memory 350, the memory page address ~~bits~~ (obtained through the encoder 340) needs to be repaired is pointed to a new remapping address ~~bits~~ in the repair memory 350, and the data are preserved. The size of repair memory 350 is determined by the length of repair address ~~bits~~ and the tolerant limit of fault pages in the slicing distribution range.

As to the multiplexer 360, it is controlled by the fault enable signal generated by the comparator array 330. As a result, it helps the system to select the original address ~~bits~~ or the remapping the address ~~bits~~ to read data.

As shown in Fig.8, the present invention includes the following steps: (1) executing memory testing to find the address ~~bits~~ and distribution of the fault in the memory chip (step 402) after the system boots (step 401); (2) if the fault distribution is in the planned range in advance, then a slicing table of fault distribution is established (step 403); (3) when the fault distribution is concentrated in a certain range, a slicing table of fault distribution is established and the address limiters 310 is posited (step 404), and the memory controller inputs the address ~~bits~~ into the address limiters 310 (step 405), and then the comparator array 330 checks to find whether the input address ~~bits~~ 380 is of different address ~~bits~~ ranges according to the slicing table of fault distribution (step 406); (4) if there are no faults, then the input address ~~bits~~ will be mapped to the SDRAM and data are read at the mapping address ~~bits~~ (step 407); (5) if there are some faults, a fault enable signal is generated and transferred to multiplexer 360, isolating the SDRAM and the original input address ~~bits~~, so that the system does not read any data at the original input address ~~bits~~ and generates a signal of fault transferred to the encoder 340 (step 408), then the encoder 340 sets up a repair address ~~bits~~ for the fault (step 409), and the repair memory 350 points to a new remapping address ~~bits~~ according to the repair address ~~bits~~ (step 410), at the same time replaces the original input address ~~bits~~ with the remapping address ~~bits~~ (step 411), data are read out at the remapping address ~~bits~~ in the SDRAM (step 412).

Add notation!!